

A Novel Gesture-based Interface for Crime Scene Investigation in Mediated Reality

Stephan Lukosch¹, Ronald Poelman¹, Oytun Akman², Pieter Jonker²

¹Delft University of Technology

Faculty of Technology, Policy and Management
Systems Engineering Section

Jaffalaan 5, 2628 BX Delft, The Netherlands

{s.g.lukosch, r.poelman, o.akman, p.p.jonker}@tudelft.nl

²Delft University of Technology

Faculty of Mechanical Engineering
BioMechanical Engineering

Mekelweg 2, 2628 CD Delft, The Netherlands

ABSTRACT

This paper introduces a novel gesture-based interface for crime scene investigation. The interface is part of a mediated reality system in which remote collaboration is supported. Requirements elicited from interviews and interactive sessions showed that our gesture-based user interface is effective in operating a 3D interface and allows the user to interact with a crime scene. We report on the design of the mediated reality system and the evaluation of the gesture-based interface. The results show that the gesture-based interface is easy to use and to learn.

Author Keywords

Mediated reality, augmented reality, gesture-based interaction.

INTRODUCTION

Crime scene investigation in the Netherlands is primarily the responsibility of the local police. For severe crimes, a national team supported by the Netherlands Forensic Institute (NFI) is called in. Initially capturing all details of a crime scene is of prime importance (so that evidence is not accidentally destroyed). NFI's Department of Digital Image Analysis uses the information collected for 3D crime scene reconstruction and analysis.

This paper reports on the results of a project between TU Delft and the NFI in which the potential of mediated [12] and augmented reality [1] for future crime scene investigation is explored. This paper focuses on a bare hand gesture interface to enable interaction and collaboration within mediated reality. To this purpose a novel mediated reality system for collaborative spatial analysis on location has been designed. This system supports collaboration between crime scene investigators (CSIs) on location and remote expert colleagues.

After discussing related work, the remainder of the paper describes the requirements for spatial analysis, the mediated reality system and in details the gesture-based interaction possibilities as well as first evaluation results.

RELATED WORK

When a severe crime is committed, the crime scene is digitalized by either photogrammetry or laser scanning methods in most western countries [6]. The most common 3D related analyses are; line of sight determination,

reconstruction of ballistic trajectories, blood pattern analysis and reconstruction of crime scenes.

To our knowledge neither mediated reality nor augmented reality is currently deployed for CSI. Augmented reality is however deployed in other domains. Fighter pilots, e.g., use a head mounted display (HMD) to display spatial information about the environment [3]. Surgeons use similar systems to overlay the patient with Magnetic Resonance Image (MRI) data during operations [4].

The GestureCam [11] system allowed a remote expert to influence the view of a novice wearing an HMD for spatial workplace collaboration. However, the technological means for superimposing more tight interaction were not available during their experiments. Many further attempts at mediated reality and augmented reality systems have been conducted over the last years; MARS, ELMO, BARS, etc. [5, 7, 9]. The pose of the system is either extracted by global positioning system (GPS) and/or inertia tracking or pre-created scene elements such as markers. Interaction with the scene is with a device or with markers.

When more natural means of interaction are considered, systems that exploit the user's hands as an interaction device are preferable since the utilization of auxiliary equipment is eliminated. Standard approaches for vision-based hand gesture tracking use hand segmentation and model fitting [13, 14]. These algorithms require either a static camera or a static background and supply 2D information which is not suitable for a 3D interface.

REQUIREMENTS FOR SUPPORTING SPATIAL ANALYSIS FOR CRIME SCENE INVESTIGATORS

Structured interviews with five international experts on the current work practices in the area of 3D crime scene reconstruction identify the challenges around the time needed for reconstruction, the necessary expertise, the varying complexity of the crime scene and the late data capture. After a brief introduction to augmented reality the following four requirements (R1-R4) were elicited in open discussion with the experts:

(R1) *Contactless augmentation alignment (no markers on the crime scene) to keep the crime scene as uncontaminated as possible.*

(R2) *Lightweight head-mounted display (HMD).*

(R3) Remote connection to and collaboration with experts to guide a novice investigator through the crime scene.

(R4) Bare hands gestures for user interface operation to have free hands to physically interact with the crime scene.

MEDIATED REALITY SYSTEM

Figure 1 depicts the mediated reality system designed on the basis of the initial requirements and the evaluation of the mock-up by practitioners. The mediated reality system consists of 4 main components: pose estimation module (PE), dense 3D map maker (MM), hand tracker (HT), networked 3D engine (NE).

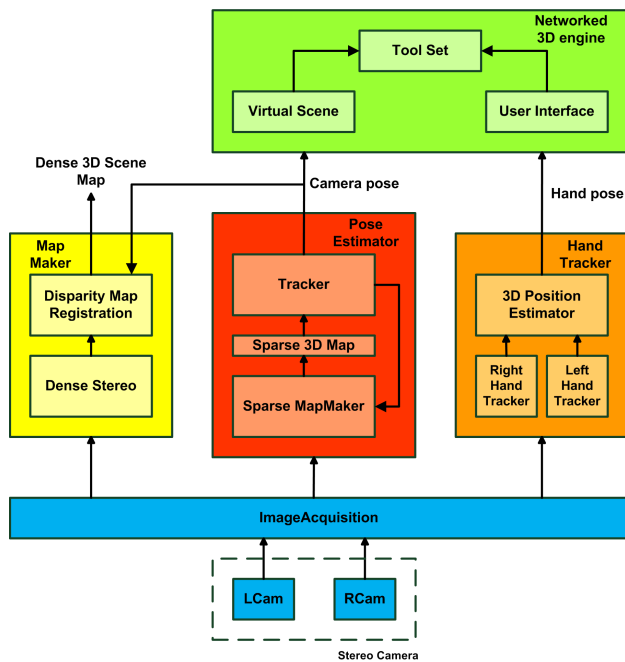


Figure 1 Overall system design

The 3D pose (position and orientation) of the user is calculated by the PE module in real time by using a heavily modified version of PTAM (Parallel Tracking and Mapping) [10] in which a single camera setup is replaced by a stereo camera setup enables 3D natural feature matching and estimation based on natural features (R1). The estimated pose is utilized by the mediated reality system to render virtual objects into the scene. The MM module uses the pose information and the stereo images to calculate a dense 3D map of the scene in near real-time.

The video see-through of a modified Carl Zeiss Cinemizer OLED (cf. Figure 2) for displaying content weighs ~170g and thus fulfills the requirement for a lightweight HMD (R2). Two Microsoft HD-5000 webcams are stripped and mounted in front of the Cinemizer providing a full stereoscopic 720p resolution pipeline.

The OGRE [8] render engine instruments connectivity between users at different locations. Thereby, it supports our requirement on a remote connection to and

collaboration with experts (R3). The server-client architecture enables all image streams, camera track pose estimations, dense maps and interactions with the scene to be uploaded to the server, the HMD wearer and remote clients to receive their information from the server. The identity of the users and their access rights, known to the system, determine user access privileges.

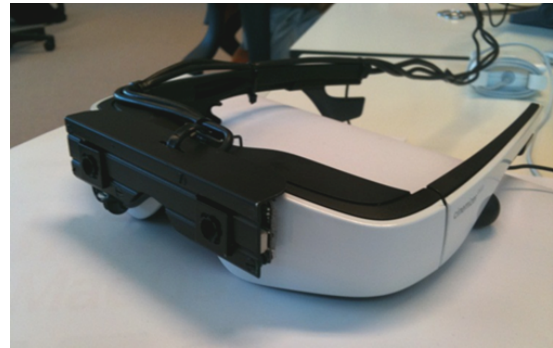


Figure 2 Head mounted display

GESTURE-BASED INTERACTION DESIGN

The hand tracker module runs independently from the other modules and scans the input image for gestures. It utilizes our stereo camera rig to detect the bare hand movements for user interface operation in 3D. The cameras are part of the HMD and a hybrid algorithm has been designed to exploit color and depth information in order to cope with changing illumination conditions, cluttered scenes and dynamic backgrounds.

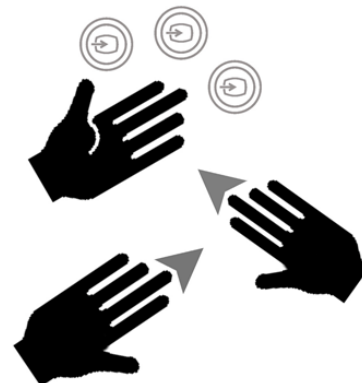


Figure 3 Types of hand postures which are recognized by the hand

Using this algorithm, we address R4 by distinguishing three types of bare-hand gestures: left hand thumb-up, left hand thumb-down, and right hand thumb-down. Figure 3 shows the gestures distinguished with the defining hand postures. A click is done by moving the left or right recognized segmented hand forward quickly, and moving it backward again. The direction of movements of the segmented hand is continuously monitored to recognize this gesture. When the pointer moves only in a forward direction, the path over which it is moving is tracked. As soon as it has moved forward, and backward more than halfway along the same path, this is registered as a click at the furthest point of the

path. If anywhere in this sequence the segmented hand deviates more than a pre-defined angle from the path, the event is not recognized as a click. In this way both small and big gestures are recognized, as long as the direction of the movement is right.

A menu surrounding the hand appears when the left hand thumb-up is detected (cf. Figure 4). The menu offers access to the following tasks for CSIs: recording the scene, placing tags, loading 3D models, bullet trajectories and placing restricted area ribbons. The menu sticks to the hand and is locked in space until the posture changes and the thumb points downwards. The right hand, as a pointing device, is used to select objects in the virtual scene.

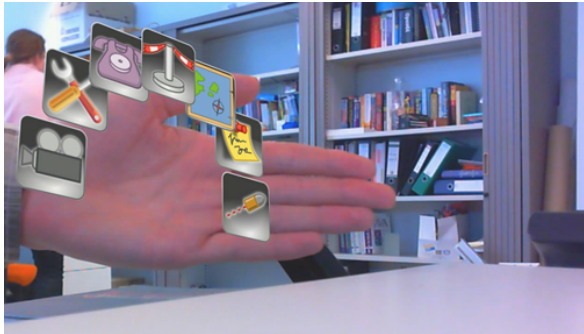


Figure 4 Graphical user interface options menu

EVALUATION SETUP

The main goal of our evaluation is to evaluate our gesture-based interface. Ten professional crime scene investigators participated in our evaluation. All received a five minutes introduction in which the test set-up and interface was explained. We performed the tests by measuring the performance of the experts in three ways. First, by logging all the hand movements, secondly by requesting feedback through a TAM-based questionnaire [2], thirdly by means of an after-action group discussion.



Figure 5 Evaluation setup

We separated the human computer interface from the AR system to evaluate it independently and ensure the same experimental setup for each participant. Instead of looking through a HMD the participants looked at a large wall projection of a stereo pre-recorded dummy crime scene (cf. Figure 5). To simulate the HMD position and detect the

gestures of the participants a stereo camera rig was mounted on a baseball cap.

We instructed the CSI experts to look at the projection, which depicts a prerecorded crime scene. Their hands can operate in the pre-recorded crime scene as if they were using the complete AR system. The experts had to conduct two tasks: 1) browsing through the options menu, in which only basic 2D GUI tasks had been loaded, and 2) tagging the crime scene by manually selecting of 3D points and placing virtual poles.

EVALUATION RESULTS

The evaluation of our tests involved the analysis of the log files that recorded the gestures of the participants, the TAM-based questionnaires, and the after-action group discussion.

The selection of the appropriate tool from the menu took most experts just one trial, for participants that did needed three trials the system failed and needed to be restarted. From a technical perspective we can conclude that 2D tool selection is working well.

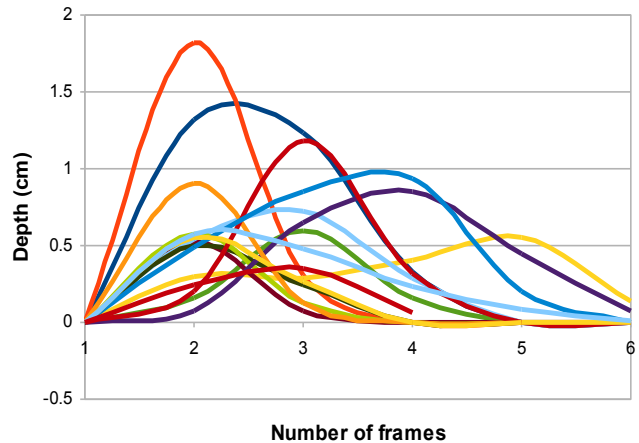


Figure 6 Gesture motion for 3D point selection of an experienced participant

For the selection of points in the real 3D scene, we monitored the depth from motion recognition to accepted command and the number of frames it took to recognize the command. Figure 6 shows the results from one experienced participant; with horizontally the number of image frames (and hence with 30 Hz implicitly the time) and vertically the observed path length of the index finger,

The higher curves show the first attempts of the participant. The lower curves show that finally after 15 attempts the user optimized its motions such that in a minimal time and minimal motion he could perform the required 3D point selection. The depth of the finger motion is less than a centimeter. The operations took just 3-6 frames which is on average 0.1 second. This shows that our algorithm is able to deal with very slight movements in a short time.

In the TAM-based questionnaire, we additionally asked each participant to provide information about age, previous

experience and field experience. By reviewing the output-graphs of the participants, the only clearly noticeable difference between the participants was their experience in working with software-based 3D models. Hence we divided the group in three classes of users; experienced users who daily use 3D models in their work, normal users who do not use it on a daily basis but are familiar with 3D models, and inexperienced users. All participants accomplished on average 15 3D placement actions to learn the 3D point selection action.

Our evaluation shows that the experienced users have one main bump in their motions, which indicates that there is not much difference between their attempts and they learned minimal quick motions to trigger the 3D point selection. The overall results of the normal users show that they are able to learn the minimal depth to trigger the action, but their motions are not as crisp as the experienced users. The inexperienced users are also able to learn the trigger motion, but they perform this slow and with abundant motion.

The TAM-based questionnaire showed similarities for both the experienced and novice users. They evaluate the gesture system as easy to use and easy to learn. They liked and felt confident in using the system, which was confirmed in the control questions. It was remarkable that the less experienced gave the highest positive remarks.

The after-action group discussion provided us with additional insights. The experienced users were comparing the interaction with the performance of their every day work, this opposed to the inexperienced users who were impressed by the system. In our log files we could also see that the experienced users invoked 3 times as much actions as the inexperienced users, which indicates that they were really testing the system. Furthermore, the participants were asked why they were making small or large gestures. We expected gestures of 3-4 centimeters and in the test most motions had been below a centimeter. Here, the experienced users claimed that small quick gestures provided them with more control and precision.

CONCLUSION

This paper presents a novel mediated reality system that builds a 3D map of the environment in real-time, allows remote users to virtually join and interact together in shared augmented space with the wearer of the HMD, and uses bare hand gestures to operate the 3D multi-touch user interface. The evaluation shows that bare-hand gesture interface allows experienced as well as lay users to interact with the mediated reality system and is easy to learn and use. In future work, we will explore how physical objects that are readily available in the environment can be used to enhance interaction, e.g. by turning a book into a menu or a coffee cup into a rotatable menu nob.

ACKNOWLEDGMENTS

Funded by Pieken in de Delta, Provincie Zuid-Holland, Gemeente Den Haag. www.csithe Hague.com

REFERENCES

1. R. Azuma, A Survey of Augmented Reality. Presence 6, Vol. 4, 355-385, 1997
2. Davis, F. D., Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 13(3), 319-340, 1989
3. E. Foxlin, Y Altshuler, L. Naimark, M. Harrington. FlightTracker: A Novel Optical/Inertial Tracker for Cockpit Enhanced Vision. Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality, 212-221, 2004
4. H. Fuchs, M. Livingston, R. Raskar. Augmented Reality Visualization for Laparoscopic Surgery. Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention, 934-943, 1998
5. T. Höllerer, S. Feiner, T. Terauchi, G. Rashid, D. Hallaway, Exploring MARS: Developing Indoor and Outdoor User Interfaces to a Mobile Augmented Reality System, Computers and Graphics, 23(6), Elsevier Publishers, 779-785, 1999
6. B. Jenkins, Laserscanning for forensic investigation, Business and technology trends in capturing and managing existing-conditions data for engineering/construction/operations, Spar Point Research LLC, Vol. 3, No. 21, 2005
7. S. Julier, Y. Baillot, M. Lanzagorta, D. Brown, L. Rosenblum. BARS: Battlefield Augmented Reality System, In NATO Symposium on Information Processing Techniques for Military Systems, 9-11, 2000
8. F. Kerger, OGRE 3D 1.7 Beginner's Guide, Packt Publishing, 2010
9. K. Kiyokawa, Y. Kurata, H. Ohno. ELMO: A Head Mount Display for Real-time Image Synthesis, ACM SIGGRAPH, 177-178, 2001
10. G. Klein, D. Murray, Parallel Tracking and Mapping for Small AR Workspaces, In Proc. International Symposium on Mixed and Augmented Reality, 225-234, 2007
11. Kuzuoka, H., T. Kosuge, GestureCam: Platform for Augmented Reality Based Collaboration International Conference on Artificial Reality and Telexistence, 1994
12. S. Mann, W. Barfield. Introduction to Mediated Reality, International Journal of Human-Computer Interaction, 205-208, 2003
13. M. Schlattman and R. Klein. Simultaneous 4 gestures 6 DOF real-time two-hand tracking without any markers. In Proc. VRST 2007, 39-42, 2007
14. R. Y. Wang and J. Popovic. Real-time hand-tracking with a color glove. ACM Transactions on Graphics, 28(3), 2009